Emotion recognition ability across different modalities: the role of language status (L1/LX), proficiency and cultural background[1]

Pernelle Lorette & Jean-Marc Dewaele

Birkbeck, University of London

Abstract

This paper considers individual differences in the Emotion Recognition Ability (ERA) of 1368 participants in different modalities. The sample consisted of 557 first language (L1) and 881 foreign language (LX) users of English from all over the world. This study investigates four independent variables, namely modality of communication, language status (L1 versus LX), proficiency, and cultural background. The dependent variable is a score reflecting ERA. Participants were asked to identify an emotion (happiness, sadness, anger, fear, surprise and disgust) portrayed by a native English-speaking actress in six short recordings – either audiovisual or audio-only – embedded in an online questionnaire. English proficiency was measured through a lexical recognition test. Statistical analyses revealed that participants were better able to recognise emotions when visual cues are available. Overall, there was no difference between L1 and LX users' ERA. However, L1 users outperformed LX users when visual cues were not available, which suggest that LX users are able to reach L1-like ERA when they can rely on a sufficient amount of cues. Participants with higher proficiency scores had significantly higher ERA scores, particularly in the audio-only condition. Asian LX users were found to score significantly lower than other LX users.

1. Introduction

Communication involves much more than purely understanding words. Next to *what* is said, one has to pay attention to *how* it is said to fully understand the essence of a conveyed message. In other words, it is crucial to decode not only linguistic, but also paralinguistic information to grasp the meaning of one's utterance. A speaker's affective orientation regarding a proposition can strongly influence the interpretation of this proposition. However, social conventions tend to discourage one's direct disclosure of emotions – especially in the case of negative emotions, leaving individuals with only indirect cues to their interlocutor's emotional state (Rintell 1984).

Every non-disabled person is able to gauge their interlocutor's emotional state to some extent – even early in life (Vailland-Molina et al. 2013). However, individual differences in this ability exist,

[1] This is the accepted version to appear in *Applied Linguistic Review* (2018)

especially in certain contexts. In this study, we focus on two of them, namely when communication occurs in a foreign language (LX)[2] (Briggs 1970; Rintell 1984) and when not all the communication channels are available (Paulmann and Pell 2011). We investigate some factors that have been shown to relate to emotion recognition ability (henceforth, ERA) in English, namely status of English (L1 or LX), proficiency in English and cultural background, and we try to gain more insight in the effect of each of these factors depending on the modality in which the communication occurs, namely audiovisual versus audio-only.

2. Literature review

2.1. Emotion: definition

Despite the long history of emotional research, there is no unanimity on the definition of emotion (e.g. Lakoff 2016; Mulligan and Scherer 2012; Pavlenko 2008). The discussion about the nature of emotions dates back to antiquity. Aristotle, for instance, analysed emotions in his work *Rhetoric* according to the beliefs, valence, actions and cognitive effects they are related to (Oatley and Jenkins 1996). Throughout the centuries, this topic kept fascinating scholars from many disciplines. The Darwinian view that emotions are the result of evolutionary selection and thus have a universal character in the human species has particularly influenced scholars working in the "basic emotion" approach, such as Paul Ekman. They conceptualise emotions as discrete, automatic, functional responses to the environment which are associated with specific physiological and behavioral reactions. (e.g. Ekman 1972, 1992). Researchers working in the appraisal framework consider emotions as originating from the cognitive evaluation of the significance of an event for the self (e.g. Scherer 1997). More recently, supporters of the integrative approach have proposed yet another perspective on emotions. They regard emotions as domain non-specific constructions of the mind which are structured around the dimensions of valence and arousal and are shaped in the course of socialisation (Barrett 2006; Russell 2003).

In brief, these frameworks define emotions in different ways, yielding different research perspectives and research questions. As applied linguists, our aim is not to confirm or disprove a specific approach, but we want to focus on the factors involved in accurate emotion recognition, as this affects communication. Because emotion recognition implies the gathering of emotional cues integrated into different channels, we adopt Keltner and Shiota's definition of emotion (2003: 89) with its focus on different channels simultaneously:

[2] We use "LX" instead of "non-native language" as opposed to "L1" instead of "native language" in order to avoid the strong connotation of inferiority related to "non-native" (see AUTHOR 2018).

> An emotion is a universal, functional reaction to an external stimulus event, temporally integrating physiological, cognitive, phenomenological, and behavioral channels to facilitate a fitness-enhancing, environment-shaping response to the current situation.

In the next sections, we will elaborate on the different channels that provide emotional information before focusing on research on individual differences in emotion recognition ability in order to highlight the gap that has motivated the present study.

2.2. Channels conveying emotional cues

Burns and Beier (1973) identified three categories of channels potentially integrating emotional information: the *verbal* – relating to the lexical content of language, the *vocal* – referring to pitch, timbre, rhythm, speaking rate, or intensity, and the *visual* channels – relating to facial expression, gesture, or body language. The cues conveyed by these channels have to be identified, sorted out depending on their relevance for the interpretation of the utterance, and interpreted accurately according to the context. This process might be more challenging under certain conditions. For instance, when not all the channels are available, the quality and / or diversity of potentially perceived information might be restricted. Moreover, in the case of cross-linguistic and/or cross-cultural communication, cues might be trickier to extract from the input or might require different interpretations according to the language and/or culture in which the communication occurs (e.g. Irvine 1982). This relates to the debate about the universal versus language/culture-specific character of emotional information conveyed via the different channels. Nowadays, researchers seem to agree that the situation is best described from a non-Manichean perspective where both are at play in the recognition of emotions (e.g. Matsumoto 2009). What still remains unclear is the ratio between universality and language/culture-specificity as well as the nature of universal features and of culture-specific features. On the one side of this debate, the staunchest advocates of universalism are researchers supporting the Ekmanian approach, as they claim that a specific set of emotions, the so-called "basic emotions" [3], are the products of biology. Those distinct emotions are assumed to be linked with distinct cues which are universally recognizable. Supporters of the integrative approach place themselves on the opposite end with an alternative account called the "Minimal Universality" (Russell 1995). It entails that the only universal features of emotions are the two primal dimensions of valence and arousal. The remaining aspects of emotion are assumed to be "constructed" by the emotion experiencer based on the context, which interpretation is affected by linguistic and/or cultural background (Russell 1991). The following sections review a number of studies contributing to this debate.

[3] The number of emotions included in the set of "basic", universally recognisable emotions has varied throughout Ekman's career (Ekman 1992, 1999, 2003). In 2011, Ekman and Cordaro (2011) claimed that emotions can only be basic, otherwise they cannot be called emotion.

2.2.1. Visual cues

Research into visual emotional cues has so far mainly focussed on the recognition of facial expression, which is assumed to be an important source of emotional information (e.g. Mesquita and Frijda 1992). Ekman and his team have conducted seminal research into cross-cultural facial ERA, mostly by using pictures of actors depicting emotional facial expression and asking participants from different cultures to pick one of the proposed labels that best describes the emotion displayed in each picture. According to his findings, emotions can be universally recognized based on their facial manifestation (e.g. Ekman et al. 1969). However, an in-group advantage has also been shown in several studies (see Elfenbein and Ambady 2002). Results indicate that participants are better at recognizing emotions displayed on the face of members of their own cultural group than those communicated by members of other cultures. Ekman and Friesen, although assuming that emotions are biologically generated, had already introduced the concept of "cultural display rules" in 1969 to account for variability in the facial expression of emotions (Ekman and Friesen 1969). They argued that these rules, acquired early in life, moderate the range of all possible human behaviours to retain only those behaviours that are appropriate in a particular culture. Crying, for instance, is an innate human behaviour expression, but cultural display rules in the Utku culture refrain Eskimos from crying - even from a young age (Briggs 1970). Despite this account for variability, studies conducted by Ekman and his team have been criticized for their methodology. Firstly, most of their findings rest on static stimuli depicting prototypical emotional expressions displayed in Caucasian faces, which might lack ecological validity (Russell 1995). Moreover, the use of forced-choice response format typically used in Ekmanian research has been questioned. According to Gendron, Roberson, van der Vyver and Barrett (2014), their study of facial ERA among American and Himba participants demonstrate that labels contained in instructions prime cross-cultural similar categorization of emotions. Without these primes, participants from different cultures categorize emotional instances in dissimilar ways.

This universality versus culture-specificity debate has not only academic relevance, but also has implications for business and economy. Tombs, Russell-Bennett and Ashkanasy (2014) researched visual ERA of 153 participants – in the role of service providers – with different cultural backgrounds – i.e. Anglo and Confucian Asian. Participants had to identify the emotional state of customers complaining in video recordings without audio. In order to minimize the effect of a forced-choice response format, participants were presented with twelve four-point Likert-type scales ranging from "nor at all feeling …" to "feeling extremely …", each of which labelled with an emotion from Richin's (1997) consumption emotion set. After the viewing of each stimulus, they had to indicate the perceived intensity for *each* of these twelve emotions. Results indicated more difficulty in the recognition of anger, happiness and shame in the case of a cultural mismatch between customers and participants, with happiness being misread in both culturally matched and mismatched dyads. Notably,

emotions expressed by Confucian Asian customers were generally more difficult to recognize than when expressed by Anglo customers. This study demonstrated firstly the great significance of ERA in multicultural companies where correct identification of emotional state can help avoid intercultural misunderstandings and lead to better service with clients. Secondly, it showed that the limitations of a forced-choice response format can be overcome while remaining within the practical boundaries of large-scale quantitative research.

2.2.2. Vocal cues

Research into vocal ERA has also greatly been concerned with investigating the respective contribution of biological and cultural factors.

Scherer, Banse, and Wallbott (2001) conducted a study in which 428 participants with different L1s and L1 cultures were asked to identify the emotion(s) expressed in meaningless multi-language sentences – i.e. meaningless strings of syllables coming from different languages – pronounced by four German actors via forced-choice response format – i.e. anger, fear, joy, sadness or neutral. Results showed an advantage of female over male participants in identifying emotions, but most interestingly, participants' country of origin turned out to have an effect on ERA, with the German participants being the best at accurately recognizing the intended emotions, followed respectively by the French-speaking Swiss – who might be familiar with German prosody since German is another official language of their country, the speakers of a Germanic language – i.e. British, Dutch and American participants, the speakers of a Romance language, – i.e. Italian, French and Spanish participants, and lastly the Indonesian participants. This finding might point to an effect of linguistic and/or cultural distance on ERA, even in the case of a forced choice between a limited set of proposed labels. However, the authors acknowledge that their finding might be a confound with judgement procedure familiarity.

Thompson and Balkwill (2006) chose to examine a homogenous emotion decoders group, which might reduce the risk of dissimilar familiarity with judgement procedure or of different conceptualization of emotion labels among the sample. The 20 participants were all L1 English users and were not fluent in any of the other languages included in the experiment. The stimuli were made of semantically-neutral utterances pronounced by English, German, Chinese, Japanese and Tagalog speakers with happy, sad, angry or scared prosody. For each stimulus, participants had to identify the expressed emotions by choosing one of the four labels joyful, sad, angry, or fearful. The participants' recognition rate was higher for emotions encoded by English speakers than for other ones. According to the authors, their findings might point to an effect of cultural distance, since the least accurately recognized emotions were the ones encoded by Japanese and Chinese speakers (respectively 59% and 54% accurate recognition). However, their findings do not comprehensively support this conclusion, since German stimuli did not yield better recognition rates than Tagalog stimuli (respectively 67.5% and 72.2% accurate recognition), as has rightly been pointed out by Pell, Monetta, Paulmann, and

Kotz (2009). Moreover, Pell and colleagues (2009) conducted a similar study with a similar methodology, which did not yield the same findings. As Thompson and Balkwill (2006), Pell and colleagues found different recognition rates for emotions encoded in the participants' L1 and emotions encoded in an unknown language. However, the recognition rates between the different unknown languages were similar.

Sauter, Eisner, Ekman, and Scott's (2010) comparison of Himba and English speakers' emotional vocalizations also highlighted another aspect that had not been controlled for in the above-mentioned studies, namely the strong bias towards negative emotions. They compared 26 English-speaking European participants with 29 Himba participants in their ability to recognize emotional nonverbal vocalizations such as scream or laughter. The vocalizations were recognized across cultures for the six "basic" emotions – note that it is unclear whether they used positively or negatively-valenced surprise vocalizations. However, some "secondary" emotions, which were interestingly all positive emotions, were not cross-culturally recognized. The authors advance an evolutionary account for this in-group advantage: since the communication of positive emotions strengthens social cohesion, it would primarily not be intended to out-group members. Similarly to Sauter and colleagues (2010), Zhu (2013) also found an in-group advantage for the perceptual ability of recognizing positive emotional prosody in Chinese and Dutch. However, the recognition rate of negative emotions such as anger or sadness was similar across cultures. Zhu argues that her findings support the hypothesis that the communication of negative emotions, as signals of danger, must be interpretable universally, irrespective of language or culture, while this might not be the case for positive emotions. Although more research is needed to confidently confirm or refute this hypothesis, Sauter and colleagues' (2010) and Zhu's (2013) findings surely underline the necessity to include more than one positive emotions in future research to avoid confounds with valence.

2.2.3. Verbal cues

Beside nonverbal cues, emotional information can also be gleaned from the actual content of utterances. However, one's personal interpretation of terms and concepts – especially abstract ones – might not always fully match the interlocutor's interpretation, particularly in multilingual or multicultural settings (Pavlenko 2008). It has been recognized that a word in language A and its "translation equivalent" in language B might activate slightly different conceptual representations such as the Spanish word "cariño" and the English closest equivalent "liking" (Altarriba 2003). Accordingly, LX users might be confronted with conceptual non-equivalence which can cause interpretation difficulties when encountering a new LX emotion concept for which they might lack the "repeated experiences" necessary to "fill" the conceptual level of the concept (Pavlenko 2008).

2.2.4. Multimodal ERA

Communication, and more specifically the communication of emotion, involves different channels simultaneously. These channels might not always provide the same amount or the same quality of information. Both in Burns and Beier's (1973) and in Collignon and colleagues' (2008) studies, participants appeared to rely more heavily on visual than on vocal cues. However, recognizers seemed to apply different strategies to focus on a particular channel depending on the quality of the information conveyed on a channel (Collignon et al. 2008), the congruence of the information conveyed via the different channels (Mehrabian and Wiener 1967), the expressed emotion (Paulmann and Pell 2011), and the recognizers' linguistic and cultural background (Riviello et al. 2011). Japanese speakers, for instance, tended to focus more on vocal cues when recognizing emotions in a multimodal setting, while Dutch speakers generally paid more attention to facial expression (Tanaka et al. 2010).

Previous research demonstrated that emotions are better recognized when conveyed in a multimodal context than when they integrate only one type of communication channel (Kreifelts et al. 2007; Collignon et al. 2008; Baenziger et al. 2009). However, the vast majority of previous studies investigating the advantage of additional channel information for emotion recognition accuracy focussed exclusively on nonverbal communication. Paulmann and Pell (2011) addressed this gap by comparing English L1 users' emotion recognition accuracy in English under six different channel conditions – i.e. visual only, vocal only, verbal only, visual-vocal, vocal-verbal, and visual-vocal-verbal. They constructed different stimuli based on video recordings of actors conveying emotions in either lexical sentences or pseudo-utterances – to eliminate any semantic cues. They presented either the raw video recordings to the participants or the extracted video or audio tracks depending on the conditions. Their statistical analyses indicated that multimodal encoding of emotions yielded better recognition rates than bimodal encoding, and that bimodal encoding yields better recognition rates than unimodal encoding. However, the considerable overlap of the standard error bars in their graph suggests that these results might only hold for this particular sample and need to be confirmed in future research.

2.3. ERA among L1 and LX users

Participants in the above-mentioned studies were all L1 users of the language(s) included in each study. It is wrong to assume that LX users behave exactly as L1 users. In her pioneering study, Rintell (1984) analysed LX users' (vocal-verbal) ERA in English. A control group of 19 L1 English users was compared to 127 LX learners of English with either Arabic, Chinese, or Spanish as L1. For each of the 11 recordings, recognizers had to choose one of the 11 labels that best characterized the emotional state of the speaker. The results revealed a main effect for status of English, with L1 participants outperforming LX participants. Moreover, the strongest effect was found for LX proficiency. The intermediate and advanced LX learners had less difficulty in identifying the emotions

in the stimuli than their less proficient peers. Not only LX users' proficiency in English, but also their L1 (culture) appeared to have an effect on the results, with the Chinese participants scoring significantly lower on ERA than the Arabic and Spanish participants. The author interpreted this finding as an effect of cultural distance.

Taking Rintell's (1984) design as a starting point, we conducted a similar study comparing English L1 and LX users' ERA, although we did not focus on vocal-verbal ERA but on visual-vocal-verbal ERA (AUTHORS 2015). Nine hundred and nineteen participants were presented with six audiovisual stimuli and had to identify the emotion conveyed by an actress by means of a forced choice. Similarly to Rintell, we found an effect of proficiency and of cultural background on ERA. However, our LX participants performed as well as our L1 participants. The difference between Rintell's and our findings might have been due to our audiovisual stimuli compared to Rintell's audio stimuli, or to the different nature of our LX users. While Rintell's LX participants were young formal learners of English enrolled in an intensive EFL course in the United States, our LX participants were older authentic LX users not necessarily enrolled in formal English classes.

Graham, Hamblin and Feldstein (2001) conducted another comparable study, which focussed on vocal ERA. One monologue was recorded eight times by several actors, each time with a different emotion conveyed in the voice. Eighty-five American-English L1 users were compared to 45 Japanese and 38 Spanish LX users of English. Just as in Rintell's (1984) study, L1 users turned out to be better at recognizing the emotions (59% correct) compared to LX users (42% for the Spanish-speaking and 38% for the Japanese-speaking participants). Moreover, the confusion patterns of the Spanish-speaking LX users were more similar to those of the control group than those of the Japanese-speaking LX users, i.e. a cultural distance effect. However, the difference between Japanese and Spanish LX users' ERA scores was not statistically significant. Contrary to Rintell (1984), Graham and colleagues (2001) did not find any effect of proficiency. The authors hypothesize that vocal ERA in an LX is only acquired after extensive exposure to the LX or if special attention is paid to vocal emotion recognition in the language classroom. However, Zhu's (2013) findings are not in line with this hypothesis. In her study, the (advanced) Dutch-speaking LX users of Chinese outperformed L1 users of Chinese in their ability to recognize emotions via vocal (prosodic) cues, although her LX participants had not received extensive exposure to Chinese in a naturalistic context.

Surprisingly, Dromey, Silveira and Sandor (2005) demonstrated that L1 users are not systematically better at vocal emotion recognition than LX users of that language. Individuals' degree of multilingualism turned out to have a stronger effect on vocal ERA than their language status. They hypothesize that their finding was due to the "additional sensitivity to certain aspects of speech" that one typically develops when learning additional languages, and that this sensitivity "carries over to native languages tasks" (Dromey et al. 2005: 356). However, as the authors point out, the number of spoken languages might have been confounded by level of education in their study, since their polyglot participants were more highly educated.

2.4. Implications for the present study

Regardless of their theoretical approach, previous studies demonstrated that a certain extent of cross-linguistic and/or cross-cultural variability appears in the ability to recognise emotions – whether encoded verbally and/or vocally and/or visually. However, there seems to be a gap in the literature about L1 and LX users' ERA in bimodal vs. multimodal settings, particularly when verbal cues remain available in both conditions. Thus, the present study will compare vocal-verbal with visual-vocal-verbal ERA and investigate the role of independent variables that have already been identified as affecting ERA, namely status of the language of users, proficiency, and cultural group.

3. Research Questions

This study aims at answering the following research questions:

1. Are emotions conveyed in English to L1 and LX users of English via the visual-vocal-verbal channels better recognizable than emotions conveyed via the vocal-verbal channels?

2.a. Are L1 users of English better able to recognize emotions conveyed in English via the visual-vocal-verbal channels compared to LX users of English?
2.b. Are L1 users of English better able to recognize emotions conveyed in English via the vocal-verbal channels compared to LX users of English only?

3.a. Are highly proficient English users better able to recognize emotions conveyed in English via the visual, the vocal, and the verbal channels compared to less proficient English users?
3.b. Are highly proficient English users better able to recognize emotions conveyed in English without visual input but only through vocal and the verbal channels compared to lower proficient English users?

4.a. Are specific cultural groups better able to recognize emotions conveyed in English via the visual, the vocal, and the verbal channels?
4.b. Are specific cultural groups better able to recognize emotions conveyed in English without visual input but only through the vocal and the verbal channels?

4. Methodology

4.1. Participants

This study combines two large datasets, with a total 1368 participants (1033 females, 335 males) who filled in either the "vocal-verbal questionnaire" (n = 449, 347 females, 102 males) or the

"visual-vocal-verbal questionnaire" (n = 919, 686 females, 233 males). Table 1 summarizes the demographics of the former, which was used in AUTHORS (2015), while Table 2 summarizes the demographics of the latter.

Table 1: Demographics of the participants who filled in the "vocal-verbal questionnaire" (*n* = 449).

	L1 users of English *(n* = 202)		**LX users of English (*n* = 247)**	
	Mean (SD)	Range	*Mean (SD)*	Range
Age	33.1 (15.9)	12-77	33.8 (12.1)	13-73
Proficiency (%) **(as measured by a lexical test–see next section)**	93.5 (7.4)	66.3-100	82.4 (12.2)	50-100

Table 2: Demographics of the participants who filled in the "visual-vocal-verbal questionnaire" (*n* = 919).

	L1 users of English (*n* = 355)		**LX users of English (*n* = 564)**	
	Mean (SD)	Range	*Mean (SD)*	Range
Age	33.3 (16.8)	11-82	30.5 (11.1)	15-70
Proficiency (%) **(as measured by a lexical test–see next section)**	94.6 (5.7)	80-100	83.4 (11.3)	45-100

The English LX users' relatively high mean proficiency score is due to the natural self-selection of the participants who had to be sufficiently proficient in English to understand the call for participation and the questionnaire, both written in English. Despite their relatively high proficiency, an independent-sample t-test revealed that the LX participants (M = 83.1, SD = 11.6) scored significantly lower on the proficiency test than the L1 participants (M = 94.2, SD = 6.4) ($t(1307.6) = 22.8$, $p < .001$).

The best represented nationality of the L1 users sample was British/Irish (*n* = 268), followed by Americans (*n* = 133). The LX users came mostly from Belgium (*n* = 161) and Slovenia (*n* = 120) but many other countries were represented, such as The Netherlands (*n* = 84) or Germany (*n* = 16). The sample was divided into nationality groups (excluding the 160 participants who reported more than one nationality) and clusters were created according to groups of nationalities, namely UK/Ireland (262 L1 users of English + 3 LX user), North-America (137 L1 users + 4 LX users), Continental Europe (23 L1 users of English + 605 LX users), Greater Middle East (6 L1 users of English + 26 LX users), and Asia (30 L1 users of English + 71 LX users of English). Participants from Australia and New Zealand, from Central and South America, and from Africa were not included because of very small sample sizes. Hereafter, these clusters will be labelled as participants' "culture". As we pointed out in AUTHORS (2015), we realize that such rough categorizations are generalizations dictated by statistical needs - as too much granularity would render statistical analysis impossible. Cultures are defined as "portable schemas of interpretation of actions and events that people have acquired through

primary socialization and which change over time as people migrate or enter into contact with people who have been socialized differently" (Kramsch 2015: 638). Although participants' original culture might have been altered due to contact with other cultures, we presume that the participants of a same cluster have been through roughly comparable primary socialization.

4.2. Instrument

We used two separate questionnaires, each of them consisting of a socio-linguistic survey, followed by six stimuli and the corresponding questions, and finally a lexical test. We call them the "visual-vocal-verbal questionnaire" – which was used in a previous study (AUTHORS 2015) – and the "vocal-verbal questionnaire" – which was specifically developed for the present study. The "vocal-verbal questionnaire" was exactly identical to the "visual-vocal-verbal questionnaire" except for one crucial distinction, namely the absence of image in the stimuli. In the following paragraphs, we describe each part of the questionnaires.

The first part of the questionnaires consisted of questions about the participants' social and linguistic background – gender, age, nationality, actual country of residence, L1 language(s) and – if applicable – LX(s), with specifications about the acquisition, use and self-rated proficiency for each language.

Depending on the questionnaire, participants were either presented with six audiovisual stimuli (in the "visual-vocal-verbal questionnaire") or with six audio recordings (in the "vocal-verbal questionnaire"). Each stimulus lasted between 30 and 55 seconds – see the Appendix for the transcriptions and the URLs of the recordings. As described in our previous study (AUTHORS 2015), the development of audiovisual stimuli was motivated by a desire to boost ecological validity by presenting a typical daily life situation in which one can simultaneously rely on visual, vocal, and verbal cues to infer the emotional state of one's interlocutor in a face-to-face conversation. Yet, communication can nowadays also occur without any visual contact between the interlocutors – for instance during a phone call. Hence the second questionnaire, with stimuli consisting only of audio recordings. In each recording, a 43-year old professional actress displayed an emotion. This actress has been born in Canada, brought up bilingually in Latvian and English and has now been living in London for several decades. English is her dominant language and her accent can best be described as English Received Pronunciation with some influences of the London accent. This lack of very strong regional accent is advantageous for our project. We choose a female actress because research has demonstrated that females' emotional state is typically better recognized than males' emotional state (e.g. Scherer et al. 2001). Each recording displayed one of the six following emotions: happiness, sadness, anger, fear, (positive) surprise, or disgust. As Sauter and colleagues (2010) and Zhu (2013) suggested that valence might affect ERA, we made sure to include a positively-valenced instance of surprise, so that happiness was not the only positive emotion included in our study. For each emotion, the actress was asked to improvise a short sketch conveying that main emotion, based either on her

own ideas or on brief example scenarios that she could choose to consult, leaving her enough freedom to ensure the authenticity of her play. In some stimuli, she conveyed more emotional cues verbally, while in other stimuli, participants had to switch their focus to nonverbal cues to gather more emotional information. This intertwining of information between the different channels corresponds to typical daily situations and made the task of the participants more challenging and intrinsically interesting. Each stimulus was presented via a Youtube-video embedded in the online questionnaire – in the "vocal-verbal questionnaire" participants listened to the actress' voice looking at a black screen. At the end of each recording, participants could click one of the six emotion labels, on "neutral emotion", or on "no idea". The number of correct identifications of the intended emotion conveyed in the 6 recordings represented the individual ERA score of each participant.

The last part of the questionnaire consisted of the English version of the LexTALE (Lemhöfer and Broersma 2012) in order to measure participants' proficiency in English. In this lexical decision test, participants are presented with 60 items one by one and have to decide whether each item is an existing (British) English word. Research has demonstrated that this test provides a reliable measure of the lexical proficiency of learners with different cultural and linguistic backgrounds and is a good indicator of overall English proficiency, at least for learners with intermediate and advanced proficiency levels (Lemhöfer and Broersma 2012).

4.3. Data collection

Participants were recruited via snowball sampling. The call for participation was launched via several social media such as Facebook and Twitter, mailing lists such as Linguistlist, and personal emails to friends and colleagues. Snowball sampling was used in order to reach as many participants as possible in as many different countries as possible. Although this method might cause the questionnaire to be spread in similar circles, we made sure to launch the questionnaire on platforms used by people with different profiles (different ages, education levels etc) in order to limit the effects of this possible drawback. The questionnaire was accessible online and there was no time limit to fill it in.

5. Results

The descriptive statistics show that most participants identified four out of six stimuli correctly (mean ERA score = 3.73, *SD* = 1.3) – see Figure 1.

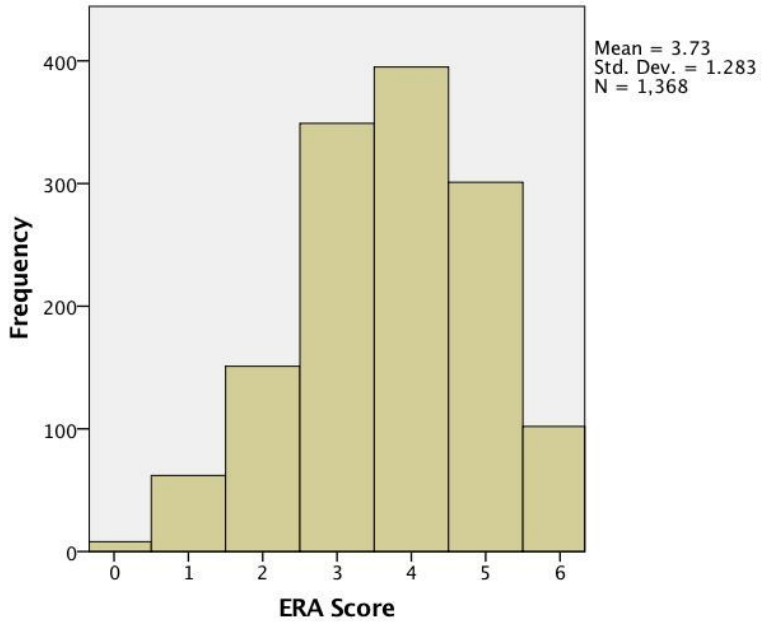

Figure 1: Bar plot showing the overall frequency of responses (n = 1386).

Bivariate correlations were used to assess the relationship between the response variable *ERA score* (henceforth *ERA*) and the other study variables *Channels*, *Status of English*, *Proficiency* and *Cultural Group.* The categorical variables *Channels*, *Status of English* and *Cultural Group* have been re-coded as dummy variables[4]. Moreover, interactions between *Channels* and the other explanatory variables as well as between *Proficiency* and *Status of English* have been examined. The variable *Proficiency* has been centered – hence renamed *Proficiency_c* – in order to reduce multicollinearity and the Bonferroni-adjusted alpha level of .003 has been adopted to correct for the number of comparisons.

Table 3: Correlations between the outcome variable ERA Score and the exploratory variables (*n* = 1368).

			ERA Score	Channels	Status of English	UK	North America	Cont. Europe	Middle East	Asia	Proficiency_c	Channels x N. America	Channels x Europe	Channels x Mid. East	Channels x Asia	Channels x Status of English	Channels x Proficiency_c	Proficiency_c x Status of English
Spearman's rho	ERA Score	Correlation Coefficient	1.000	.432**	-.001	.003	.016	.105**	-.089**	-.129**	.112**	.083**	.292**	-.007	-.064*	.271**	.098**	.061*
		Sig. (2-tailed)	.	.000	.968	.920	.547	.000	.001	.000	.000	.002	.000	.784	.019	.000	.000	.025
		N	1368	1368	1368	1368	1368	1368	1368	1368	1368	1368	1368	1368	1368	1368	1368	1368

Table 3 shows that *ERA* was positively correlated with *Channels (ρ= .432, p < .001),* indicating that participants were significantly better at recognising emotions when visual, vocal, and verbal cues were simultaneously available (mean *ERA* = 4.1) than when they could only rely on vocal and verbal cues (mean *ERA* = 2.9). Regarding cultural groups, *Continental Europe* (henceforth

[4] Note that SPSS automatically computes Point-biserial correlation coefficients when a dichotomous term is involved in a correlation calculation.

Europe, mean *ERA* = 3.9, ρ = .105, p < .001) was positively correlated with *ERA,* whereas *Middle East (*mean *ERA* = 2.9, ρ = -.089, p < .001), and *Asia (*mean *ERA* = 3.2, ρ = -.129, p < .001) were negatively correlated with *ERA*. This suggests that European participants were better able to recognise emotions when compared to the rest of the sample, and that participants from the Middle East and from Asia had more difficulties in recognising emotions in English. The Spearman's rhos also indicate a relationship between *Proficiency_c* and *ERA* (ρ = .112, p < .001), demonstrating that the more proficient a speaker is in English, the better (s)he is able to recognise emotions in this language. Turning to the interactions, *Channels* appeared to be an important moderator, as *Channels* × *North America* (ρ =.083, p <.002), *Channels* × *Europe* (ρ = .292, p <.000), and *Channels* × *Status of English* (ρ = .271, p < .001) were significantly correlated with *ERA*.

To interpret the latter interaction, Mann-Whitney U tests have been run to compare L1 and LX users separately for vocal-verbal and visual-vocal-verbal ERA. Figure 2 shows that L1 users are better than LX users at recognising emotions in English when only vocal and verbal cues are available (U = 20701, p < .001, r = -.15, n = 449). However, this difference disappears when visual cues are also available (U = 96424, p > .05, n = 919), meaning that our LX participants can recognise emotions as well as the L1 participants when they can rely simultaneously on visual, vocal and verbal cues.

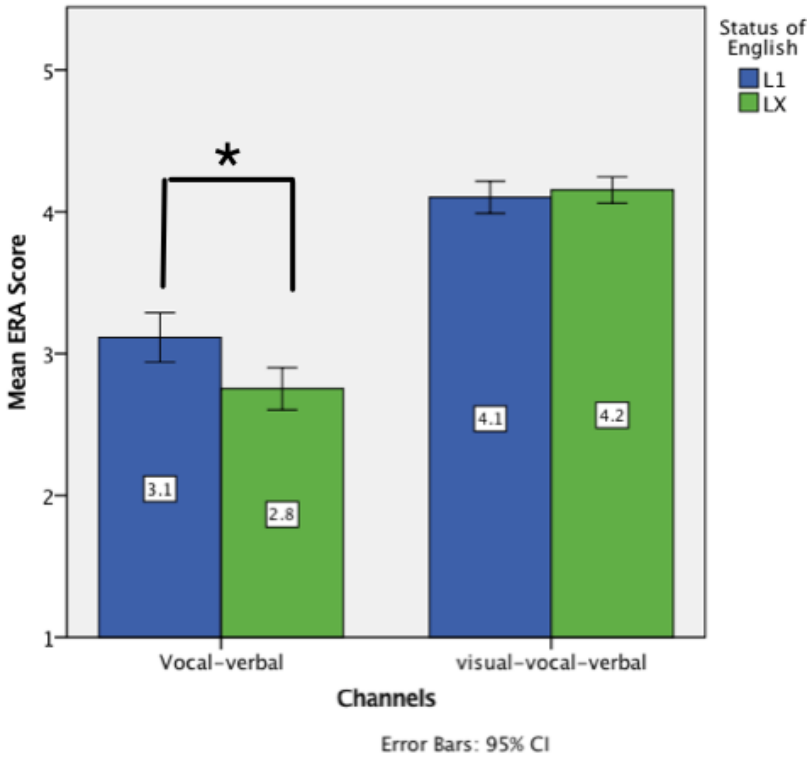

Figure 2: Barplot showing the difference between L1 and LX speakers' ERA when vocal-verbal (n = 447) and when visual-vocal-verbal cues (n = 919) are available (* p < .05)

The two interactions involving cultural groups – i.e. *Channels* × *North-America* and *Channels* × *Europe* – are less straightforward to interpret. Figure 3 shows a trend towards higher ERA scores of North-American compared to other cultural groups in vocal-verbal ERA. Conversely, it is the

European that seem to outperform the other cultural groups when visual, vocal, and verbal cues are available. However, the error bars show that these trends need to be confirmed in further analyses.

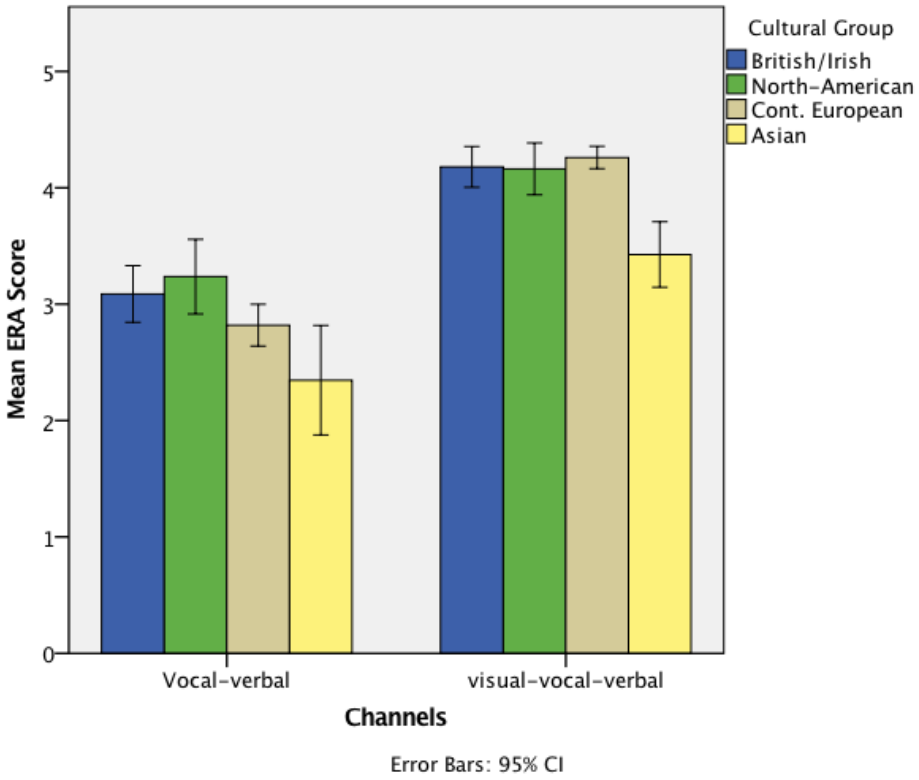

Figure 3: Mean ERA scores for vocal-verbal vs. visual-vocal-verbal channels, grouped by cultural group (n = 1368)

In order to determine whether all these effects hold when the other independent variables are controlled for, the above-mentioned correlated variables were fed into a two-stage hierarchical linear regression model, assessing which factors predict the outcome variable *ERA*. The main effects were entered at stage one and the interactions were entered at stage two. Due to heteroscedasticity and non-normality of the residuals, the bootstrapping approach has been used with 95% Biased-corrected accelerated confidence intervals (Bca CI) based on 1000 bootstrap samples.

At stage one, the five-predictor regression model was able to account for 23% of the variance in *ERA* ($F(5, 1362) = 82.18$, $p < .001$). *Channels* ($p < .001$, 95% CI [1.074, 1.329]), *Asia* ($p < .001$, 95% CI [-.927, -.357]) and *Proficiency_c* ($p < .001$, 95% CI [.007, .019) appeared to be significant predictors of *ERA* – see Table 4. At stage two, all interaction variables correlating with *ERA* were added to the model, resulting in a significant but rather limited increase of .6% of explained variance ($F(4, 1358) = 3.04$, $p < .017$), with the full model explaining 23.6% of the variance in ERA ($F(9,1358) = 47.8$, $p < .001$). However, this significant R^2 increase did not seem to hold in the bootstrap model: among the variables added at stage two, only *Channels* × *Europe* reached significance in the initial model ($p < .036$), but barely reached marginal significance in the bootstrap model ($p = .054$, 95% CI [.017, .644]).

Table 4: Summary of the hierarchical regression model of ERA scores with 5 main effect predictors (step 1) and 4 interaction predictors (step 2) (n = 1368).

Model		B	Bootstrap[a]				
			Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval	
						Lower	Upper
1	(Constant)	2.923	-.002	.065	.001	2.792	3.051
	Channels	1.199	-.001	.067	.001	1.074	1.329
	Cont. Europe	.122	.004	.066	.071	-.014	.273
	Middle East	-.310	-.004	.245	.190	-.735	.164
	Asia	-.594	.001	.127	.001	-.827	-.357
	Proficiency_c	.013	.000	.003	.001	.007	.019
2	(Constant)	3.009	-.002	.075	.001	2.865	3.154
	Channels	1.050	.000	.103	.001	.831	1.248
	Cont. Europe	-.106	.005	.122	.376	-.365	.143
	Middle East	-.344	-.004	.246	.153	-.768	.118
	Asia	-.562	-.001	.133	.001	-.809	-.305
	Proficiency_c	.020	1.522E-5	.005	.001	.010	.029
	Channels x Status of English	.013	.001	.128	.933	-.226	.262
	Channels x N. America	.033	-.004	.123	.784	-.215	.267
	Channels x Cont. Europe	.329	-.003	.164	.054	.017	.644
	Channels x Proficiency_c	-.010	.000	.006	.114	-.022	.003

In summary, based on the initial (non-bootstrapped) standardized coefficients, *channels* (*Beta* = .439) seems to be the best predictor of *ERA*, followed by *proficiency_c* (*Beta* = .173) and *Asia* (-.115). In order to further understand the mechanisms influencing ERA, correlation and regression analyses similar to the above-mentioned ones were conducted separately for visual-vocal-verbal ERA and for vocal-verbal ERA. Regarding vocal-verbal ERA, the only significant predictor in our hierarchical regression model ($F(5, 443) = 7.03$, $p < .001$, $R^2 = .06$) appeared to be *Proficiency_c* ($p = .001$, 95% CI [.026, .071]). For visual-vocal-verbal ERA, only *Asian* ($p = .001$, 95% CI [-.983, -.402]) appeared to significantly predict *ERA* in our hierarchical regression model ($F(2, 916) = 18.94$, $p < .001$, $R^2 = .04$). Post-hoc Mann-Whitney U tests confirmed the significant difference between Asian and British/Irish participants ($U = 4046$, $p < .001$, $r = -.28$), North-American ($U = 2126$, $p < .001$, $r = -.30$) and European ($U = 10658$, $p < .001$, $r = -.24$).

6. Discussion

The findings of this study revealed that the modality in which emotions are conveyed is a significant predictor of ERA. Emotion recognition is significantly less accurate when emotions only integrate the vocal and verbal channels compared to when additional visual cues are available. The present study is, to our knowledge, the only one in which verbal cues remained identical in the different conditions, which allowed us to conclude unambiguously that an extra channel boosts emotion recognition. As such it confirms and expands previous findings (e.g. Collignon et al. 2008; de Gelder and Vroomen 2000; Kreifelts et al. 2007; Paulmann and Pell 2011).

Regarding the comparison of L1 and LX users, correlations indicated no overall difference between L1 and LX users' ability to recognise emotions. However, when controlling for the modality in which emotions are conveyed, L1 users appear more accurate than LX users at recognising emotions conveyed without visual cues – which is consistent with Rintell's (1984) and Graham and colleagues' (2001) findings. This suggests that recognising emotions in a LX context might be more challenging than in a L1 context, but that additional cues might help LX users to reach L1 users' recognition accuracy rate. Moreover, this finding might point to the more universal character of visual displays of emotion compared to vocal displays, although further research is needed to verify this speculation.

Furthermore, the present study revealed a relationship between ERA and proficiency, which confirms Rintell's (1984), Graham and colleagues' (2001) and our own findings (AUTHORS 2015). Proficiency is an important predictor of ERA. The more linguistically proficient an individual is – regardless of whether that person is a L1 or a LX speaker – the more accurate (s)he will be at recognising (basic) emotions. An interesting finding is that when we analysed the data separately for visual-vocal-verbal and for vocal-verbal ERA, proficiency only appeared to be a significant predictor of vocal-verbal ERA, but not of visual-vocal-verbal ERA. This might suggest that low-proficient English speakers are able to compensate for their lower lexical knowledge by relying on visual cues, but lack the necessary cues to do so when only vocal cues are available beside the verbal ones.

Our results confirmed the effect of cultural background on ERA found in previous research. Particularly, Asian cultural background was the third significant predictor of ERA, with an Asian disadvantage in recognizing emotions in English (Rintell 1984, Tombs et al. 2014, AUTHORS 2015). The negative relationship between ERA and Asian cultural background has been found across both modality conditions investigated in this study. These findings chime with Zhu's (2013) and Tanaka and colleagues' (2010) study about respectively Chinese vs. Dutch and Japanese vs. Dutch individuals' ERA. These findings might result from differences in affective socialization between Eastern and Western cultures, as proposed by Wang and Ross (2005).

7. Limitations and perspectives for further research

We are aware of a number of limitations in the research design. First of all, the use of a professional actress for all the self-constructed stimuli has a number of drawbacks. Since the emotions were acted, one might argue that they do not totally reflect natural, unconsciously encoded emotions that individuals are confronted with in their daily live. Moreover, the patterns found in the study are necessarily linked to the actress' unique portrayal of emotions, and might therefore not be generalizable to other situations. This limitation could be avoided by using several actors for each emotion in order to limit the impact of one personal portrayal of emotions on the ERA scores.

Secondly, the emotions conveyed in the stimuli were not entirely "pure" - if such a thing exists - but several emotions were at play in each stimulus. However, we believe that this reflects daily live situations, thus strengthening the ecological validity of the study. We also anticipated this limitation by asking the participants to identify the "main" emotion conveyed in each stimulus, acknowledging that other emotions might be present to a smaller extent in each stimulus.

Thirdly, in some recordings, the emotional event itself was acted in the present whereas in other recordings, a past emotional event was reported. However, the retelling of a past emotional event often reactivates the emotion in the storyteller. The actress was particularly careful to report the event as if she was re-experiencing the emotion at play. Therefore, we are pretty confident that this limitation did not have much impact on the findings.

Fourthly, our attempt to cluster our participants according to their cultural background was only a rough categorization that needs to be replicated and refined in further research. Grouping participants always implies some form of (over)generalisation, and especially nationalities do not always reflect one's "culture". It is clear from the literature that it is delicate to investigate an effect of "culture" on any linguistic aspect, since: a) culture is a very broad concept and can be understood in many different ways (see Gladkova 2014); and b) culture and language are intertwined and therefore very difficult to set apart (see AUTHOR, 2015; Robinson and Altarriba 2014). A larger sample, consisting of homogeneous categories with comparable numbers of participants could yield richer and more fine-grained results.

Lastly, in it important to keep in mind that our findings only apply to so-called "basic" emotions displayed in a rather "prototypical" context – with no or very limited amount of irony or incongruence between the different channels. Further research should attempt to include other emotions in the stimuli in order to investigate whether language status, proficiency and cultural background have stronger effects in ambiguous or incongruent input.

8. Conclusion

The finding that accurate emotion recognition depends on the number of channels in the input (vocal, verbal and visual) and that the absence of the visual channel impairs recognition by L1 and LX users confirms previous research using a better design. The most original finding is that less proficient LX users suffer most from the absence of visual input. Finally, linguistic and/or cultural distance affects English emotion recognition in both audio-only and audio-visual stimuli, with Asian participants experiencing more difficulties.

9. Bibliography

Baenziger, Tanja, Didier Grandjean & Klaus R. Scherer. 2009. Emotion recognition from expressions in face, voice, and body. The Multimodal Emotion Recognition Test (MERT). *Emotion* 9(5). 691–704.

Barrett, Lisa F. 2006. Solving the Emotion Paradox: Categorization and the Experience of Emotion. *Personality and Social Psychology Review* 10(1). 20–46.

Jean L. Briggs. 1970. *Never in anger: Portrait of an Eskimo family*. Cambridge, MA: Harvard University Press.

Burns, Kenton L. & Ernst G. Beier. 1973. Significance of vocal and visual channels in the decoding of emotional meaning. *The Journal of Communication* 23. 118–130.

Collignon, Olivier, Simon Girard, Frederic Gosselin, Sylvain Roy, Dave Saint-Amour, Maryse Lassonde & Franco Lepore. 2008. Audio-visual integration of emotion expression. *Brain Research* 1242. 126–135.

De Gelder, Beatrice & Jean Vroomen. The perception of emotions by ear and by eye. *Cognition and Emotion* 14(3). 289-311.

Dromey, Christopher, Jose Silveira & Paul Sandor. 2005. Recognition of affective prosody by speakers of English as a first or foreign language. *Speech Communication* 47. 351–359.

Ekman, Paul. 1972. Universals and cultural differences in facial expression of emotion. In James (ed.), *Nebraska Symposium on Motivation* (Volume 19), 207-283. Lincoln: University of Nebraska Press.

Ekman, Paul. 1992. An argument for basic emotions. *Cognition and Emotion* 6. 169–200.

Ekman, Paul. 1999. Basic Emotions. In Tim Dalgleish & Michaall Power (eds.), *Handbook of Cognition and Emotion*, 45–60. Sussex, UK: John Wiley.

Ekman, Paul. 2003. Sixteen enjoyable emotions. *Emotion Researcher* 18. 6–7.

Ekman, Paul & Daniel Cordaro. 2011. What is meant by calling emotions basic. *Emotion Review* 3(4). 354–370.

Ekman, Paul & Wallace V. Friesen. 1969. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica* 1. 49–98.

Elfenbein, Hillary A. & Nalini Ambady. 2002. On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin* 128(2). 203–235.

Gendron, Maria, Debi Roberson, Jacoba M. van der Vyver & Lisa F. Barrett. 2014. Perceptions of emotion from facial expressions are not culturally universal: Evidence from a remote culture. *Emotion* 14(2). 251–262.

Graham, Ray C., Arien W. Hamblin, & Stanley Feldstein. 2001. Recognition of emotion in English voices by speakers of Japanese, Spanish and English. *International Review of Applied Linguistics in Language Teaching* 39. 19–37.

Gladkova, Anna. 2014. Ethnosyntax. In Farzad Sharifian (ed.). *The Routledge Handbook of Language and Culture*, 52-88. London, New York: Routledge.

Irvine, Judith T. 1982. Language and affect: Some cross-cultural issues. In Heidi Byrnes (ed.), *Contemporary Perceptions of Language: Interdisciplinary Dimensions,* 31–47. Washington, D.C: GU Press.

Keltner, Dacher & Michelle N. Shiota. 2003. New displays and new emotions: A commentary on Rozin and Cohen (2003). *Emotion* 3. 86–91.

Kramsch, Claire. 2015. Language and culture in foreign language learning. In Farzad Sharifian (ed.), *The Routledge handbook of language and culture*, 634–644. Oxford: Routledge.

Kreifelts, Benjamin, Thomas Ethofer, Wolfgang Grodd, Michael Erb & Dirk Wildgruber. 2007. Audiovisual integration of emotional signals in voice and face: An event-related fMRI study. *Neuroimage* 37. 1445–1456.

Lakoff, George. 2016. Language and emotion. *Emotion Review* 8(3). 269-273.

Lemhöfer, Kristin & Mirjam Broersma. 2012. Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods* 44(2). 325–343.

Matsumoto, David. (2009). *Facial expressions are not universal*. https://digest.bps.org.uk/2009/08/17/facial-emotional-expressions-are-not-universal/ (accessed 09 November 2015)

Mehrabian, Albert & Morton Wiener. 1967. Decoding of inconsistent communications. *Journal of Personality and Social Psychology* 6. 108–114.

Mesquita, Batja & Nico H. Frijda. 1992. Cultural variations in emotions: A review. *Psychological Bulletin* 112(2). 179–204.

Mulligan, Kevin & Klaus R. Scherer. 2012. Toward a Working Definition of Emotion. *Emotion Review* 4(4). 345-357.

Oatley, Keith & Jennifer M. Jenkins. 1996. *Understanding Emotions*. Oxford, UK: Blackwell Publishing.

Paulmann, Silke & Marc D. Pell. 2011. Is there an advantage for recognizing multi-modal emotional stimuli? *Motivation and Emotion* 35. 192–201.

Pavlenko, Aneta. 2008. Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition* 11(2). 147–164.

Pell, Marc D., Laura Monetta, Silke Paulmann & Sonja A. Kotz. 2009. Recognizing emotions in a foreign language. *Journal of Nonverbal Behavior* 33. 107–120.

Rintell, Ellen. 1984. But how did you FEEL about that? The learner's perception of emotion in speech. *Applied Linguistics* 5(3). 255–264.

Riviello, Maria Teresa, Anna Esposito, Mohamed Chetouani & David Cohen. 2011. Inferring emotional information from vocal and visual cues: A cross-cultural comparison. *2nd IEEE International Conference on Cognitive Infocommunications (CogInfoCom*). 1–4.

Robinson, Crystal J. & Jeanette Altarrriba. 2014. Culture and Language Processing. In Farzad Sharifian (ed.), *The Routledge Handbook of Language and Culture,* 355-381. London, New York: Routledge.

Russell, James A. 1991. Culture and the categorization of emotion. *Psychological Bulletin* 110. 426–450.

Russell, James A. 1995. Facial expressions of emotion: What lies beyond minimal universality? *Psychological Bulletin* 118. 379–391.

Russell, James A. 2003. Core affect and the psychological construction of emotion. *Psychological Review* 110. 145–172.

Sauter, Disa A., Frank Eisner, Paul Ekman & Sophie K. Scott. *2010. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proceedings of the National Academy of Sciences of the United States of America* 107(6). 2408–2412.

Scherer, Klaus R. 1997. The role of culture in emotion-antecedent appraisal. *Journal of Personality and Social Psychology* 73(5). 902–922.

Scherer, Klaus R., Reiner Banse & Harad G. Wallbott. 2001. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology* 32(1). 76–92.

Tanaka, Akihiro, Ai Koizumi, Hisato Imai, Saori Hiramatsu, Eriko Hiramoto & Beatrice de Gelder. 2010. I feel your voice. Cultural differences in the multisensory perception of emotion. *Psychological Sciences* 21. 1259–62.

Thompson, William F. & Laura-Lee Balkwill. 2006. Decoding speech prosody in five languages. *Semiotica* 158. 407–424.

Tombs, Alastair G., Rebekah Russell-Bennett & Neal M. Ashkanasy. 2014. Recognising emotional expressions of complaining customers: A cross-cultural study. *European Journal of Marketing* 48(7/8). 1354–1374.

Vaillant-Molina, Mariana, Lorraine E. Bahrick & Ross Flom. 2013. Young infants match facial and vocal emotional expressions of other infants. *Infancy* 18. E97–E111.

Wang, Qi & Michael Ross. 2005. What we remember and what we tell: the effects of culture and self-priming on memory representations and narratives. *Memory* 13(6). 594–606.

Zhu, Yinyin. 2013. *Expression and recognition of emotion in native and foreign speech: The case of Mandarin and Dutch*. PhD dissertation, Leiden University.

Appendix

The stimuli can be consulted on the following links. They have been transcribed below, in order of apparition in the survey.

1) Disgust: http://www.youtube.com/embed/rH6evcth9Vs (visual-vocal-verbal) and https://www.youtube.com/watch?v=UUU0vreMgj8 (vocal-verbal)

So, Jerry, you wanted to discuss the proposal that I put together for the two separate groups. You, you, you've got something... Kind of... No, no, it's not... It's sort of there. No, it's still there. It's now dripping down a little bit. Maybe if you use a napkin somewhere that you could wipe it with.

2) Anger: http://www.youtube.com/embed/8VcoNbk3HVE (visual-vocal-verbal) and https://www.youtube.com/watch?v=5XDGUAUh54A (vocal-verbal)

Yesterday, I went to see my mother-in-law. It was actually her birthday the day before yesterday, but I couldn't go because I had a business meeting. And I bought her a very nice bunch of flowers. Very nice. And when I got there, she said:"What is this about?". And I said: "Well, it is your birthday, Maria. Happy birthday!" And she said: "It's not my birthday, it was my birthday yesterday." So anyway, I really hope she liked the flowers.

3) Happiness: http://www.youtube.com/embed/x1S3IzTmf6A (visual-vocal-verbal) and https://www.youtube.com/watch?v=7SlWJk5N-iQ (vocal-verbal)

So, I went to my Pilates class after a really long time of absence of a few weeks, which you start to really notice if you haven't been. But the teacher is absolutely amazing. What she's really into is torturing us, basically. And she, she wants you to work really really hard. And she says: "Oh, when I'm coming in a... You know, if I am in a bad mood, if I see you there and I can hear you groaning a little bit, and gasping and running out of breath, then I think "Brilliant, I'm really getting them to do some good work".

4) Fear: http://www.youtube.com/embed/T_5uBEYC8Wc (visual-vocal-verbal) and https://www.youtube.com/watch?v=WAWeL1UQhac (vocal-verbal)

So, I've got quite bad back pain, and it's been like that for about three weeks. It's really on my right side. And I suppose what I want to know is what... what it... you know, because I've tried doing some stretching but they haven't... haven't really worked at all. And I just kind of wondered whether you could tell me if you could exclude some things that it could be. It's just that I know that one of the indications is some kind of... I know this sounds stupid but... some indications of... And I know I'm probably fine but... some indications of.. of... of... certain kinds of cancers can be... to do with back pain. And that's kind of when... I don't know if you can kind of just eliminate it. That would be really helpful.

5) Surprise: http://www.youtube.com/embed/rHuCJ6rojzE (visual-vocal-verbal) and https://www.youtube.com/watch?v=C8cO-UYimcY (vocal-verbal)

So this is like a really beautiful restaurant. It's just really really nice, and... I just, you know, kind of... Oh my god! Really? Yes, Okay!

6) Sadness: http://www.youtube.com/embed/B-k3ivqrVDw (visual-vocal-verbal) and https://www.youtube.com/watch?v=rwDKHSamKp8 (vocal-verbal)

So, yesterday, I went to see my mother in law. It was actually her birthday the day before that and I actually couldn't go. I was, you know, away working. So, I went the following day. And I bought her some flowers, and gave her the bouquet. And she was asking me why I bought her some flowers. And I said: "Well, because it is your birthday, Maria." And she said: "No, it isn't, it was my birthday the day before." So, yeah, well anyway, I really hope she liked the flowers.